Signal Waveform Estimation in the Presence of Uncertainties About the Steering Vector

Olivier Besson, *Member, IEEE*, Andrei A. Monakov, and Christophe Chalus

***Abstract*—We consider the problem of signal waveform estimation using an array of sensors where there exist uncertainties about the steering vector of interest. This problem occurs in many situations, including arrays undergoing deformations, uncalibrated arrays, scattering around the source, etc. In this paper, we assume that some statistical knowledge about the variations of the steering vector is available. Within this framework, two approaches are proposed, depending on whether the signal is assumed to be deterministic or random. In the former case, the maximum likelihood (ML) estimator is derived. It is shown that it amounts to a beamforming-like processing of the observations, and an iterative algorithm is presented to obtain the ML weight vector. For random signals, a Bayesian approach is advocated, and we successively derive an (approximate) minimum mean-square error estimator and maximum *a posteriori* estimators. Numerical examples are provided to illustrate the performances of the estimators.**

***Index Terms*—Array processing, beamforming, signal waveform estimation, steering vector errors.**

I. Introduction

In future generations of satellite communication systems, the payload is likely to include arrays of sensors together with onboard digital beamforming capabilities in order to improve coverage and capacity and to enhance the flexibility of information processing. Such systems use a multibeam concept in which multiple beams ("spots") are formed, each of them covering a region whose diameter is approximately a few hundreds of kilometers, which corresponds to a $1°$–$2°$ beamwidth, as seen from a geostationary satellite; see, e.g., [1] for the description of such a concept. Since there is already some experience of using them and because they turn out to be the most promising solution in reception, a focal array fed reflector (FAFR) remains an interesting solution within this framework. A FAFR naturally concentrates energy impinging from each spot on a small number of feeds in its focal plane. Therefore, only a few feeds need to be controlled in order to form a spot, which simplifies the beamforming operation. For instance, one can typically have hundreds of feeds in the array, and only seven to 12 of them are used for each spot.

Manuscript received April 10, 2003; revised September 25, 2003. The work of O. Besson and C. Chalus was supported in part by Alcatel Space under Contract A62400. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alexi Gorokhov.

O. Besson is with the Department of Avionics and Systems, ENSICA, Toulouse, France (e-mail: besson@ensica.fr).

A. A. Monakov is with St. Petersburg State University of Aerospace Instrumentation, Saint Petersburg, Russia (e-mail: and@aanet.ru).

C. Chalus is with the Cooperative Research Laboratory for Aeronautics and Space Telecommunications (TéSA), Toulouse, France (e-mail: christophe.chalus@tesa.prd.fr).

Digital Object Identifier 10.1109/TSP.2004.831917

In order to obtain high gains and sufficient isolation between spots, large apertures may be required (depending on the frequency band in which the satellite operates), which in turn implies an increased weight. Since weight is a major concern in satellite systems, some constraints may be relaxed, e.g., on the system that ensures rigidity of the paraboloid. Then, the paraboloid undergoes deformation, and the beamformer weights that have been calculated based on a nominal position of the array are no longer capable of providing the required level of performance. In particular, signal waveform estimation may be degraded. Accordingly, thermal and gravitational effects as well as surface errors give rise to similar effects. Consequently, it becomes important to compensate for these effects and to recover a beamformer that still maintains a specified performance under these conditions. This is the main problem we address in this paper. A few specificities of the above-mentioned problem are worthy of mention.

- In a FAFR, the deformations or surface errors mainly affect the main parabola, and therefore, the feeds are no longer exactly located in the focal plane.
- In contrast to a direct radiating array (DRA), there does not exist, for a FAFR, any simple, analytical model for the steering vector associated with a source impinging on the array as a function of its direction of arrival (DOA) and the positions of the antenna elements. Only electromagnetism simulation tools are able to recover the spatial signature. In other words, the usual model $\boldsymbol{a}(\theta, \boldsymbol{\rho})$, where θ denotes the DOA and $\boldsymbol{\rho}$ is the vector of sensor locations, cannot be used.
- Models for the possible deformations of the FAFR are available [2], [3]. These models, along with electromagnetism numerical tools, are able to characterize the variations of the spatial signatures.

To summarize, we consider herein the problem of signal waveform estimation, where, on one hand, there exist uncertainties about the spatial signature of interest, but on the other hand, some *a priori* information about the variations of the steering vectors is available. Usually, in order to cope with this problem, two main approaches can be advocated. The first consists in designing a robust beamformer, i.e., a beamformer whose performance does not dramatically deteriorate in the presence of uncertainties. The second approach consists of estimating the spatial signature (including those terms that affect it such as gain and phases, etc) and then designing a beamformer based on the estimated steering vector. This second approach is often referred to as calibration.

Robust adaptive beamforming has attracted a large amount of attention in recent decades; see, e.g., [4, ch. 6] and [5] for comprehensive overviews. The most widely used method [due to its simplicity and effectiveness] is diagonal loading [6]. Diagonal loading can either be viewed as a means to "equalize" the least significant eigenvalues of the sample covariance matrix or to constrain the array gain. The latter interpretation also suggests that it can be effective when dealing with uncalibrated arrays (see [4, ch. 6]). Steering vector mismatch is often treated via the introduction of linear constraints. The latter include directional constraints, derivative constraints, quadratic constraints on the norm of the weight vector [7], and soft constrained beamformers [8]–[10]. Accordingly, null broadening techniques [11], [12] or covariance matrix tapering [13], [14] can be employed. Finally, an effective approach was proposed in [15] (with a generalization in [16]), which consists of projecting the presumed steering vector onto the subspace where it is expected to lie. Generally, a combination of these techniques is required, depending on the application. The latest developments in this area [17]–[19] concern worst-case approaches whose idea is to ensure that the response of the beamformer be above some level for all steering vectors whose distance to the presumed steering vector is less than a certain distance. Essentially, these methods amount to generalized (i.e., not necessarily diagonal) loading of the sample covariance matrix. Most of the above-mentioned approaches make no use of some *a priori* knowledge in some statistical form, except for a few works. Only recently, Bell and her co-authors have advocated in [20] a Bayesian approach to address the case where the direction of arrival is assumed to be a discrete random variable with a known probability density function (pdf). This allows the design of a beamformer that is robust to uncertainties in DOA within a neat statistical framework.

The second way to tackle the uncertainty about the steering vector is to use an explicit model for the steering vector that is usually in terms of the DOA and some nuisance parameters such as gain and phases, sensor locations, etc. The beamformer is then designed based on their estimates. To obtain the latter, two different approaches can be taken; either consider the nuisance parameters as deterministic (see, e.g., [21] and [22] or as random [23]–[26]). In the latter category, the nuisance parameters are assigned an *a priori* pdf, and for instance, MAP approaches are considered in [23] and [25]. However, these references consider a model for the steering vector of the type $\boldsymbol{a}(\theta, \boldsymbol{\rho})$, where $\boldsymbol{\rho}$ is a random vector that consists, e.g., of the sensor gains and phases or locations. Unfortunately, as stated in the introduction, such a model is not applicable for an FAFR, and hence, the corresponding methods cannot be applied. This is why we turn to another model where the steering vector is random, but we do not try to explicitly parameterize it in terms of some specific parameters. The advantage of doing so is that the model is general, and the analysis to be presented is suitable for many problems. In other words, we do not try to identify the reason why the steering vector is random; indeed, in most applications, this uncertainty can be due to many reasons, including scattering around the source, propagation through an inhomogeneous medium, or uncalibrated or displaced sensors.

The paper is organized as follows. In Section II, we present the model used through the paper. Section III considers maximum likelihood signal waveform estimation, while Bayesian approaches are proposed in Section IV. Numerical simulations are reported in Section V, and our conclusions are drawn in Section VI.

II. Data Model

Let us consider an m-element array. Assuming narrowband processing, the output of the array can be written as

$$\boldsymbol{x}_t = \boldsymbol{a}_s s_t + \boldsymbol{n}_t \quad t = 1, \dots, N \tag{1}$$

where $\boldsymbol{a}_s$ is the steering vector of interest, and s_t is the corresponding emitted signal. $\boldsymbol{n}_t$ denotes the noise contribution (possibly including interferences). In the sequel, we make the following assumptions.

- We assume that $\boldsymbol{a}_s$ is drawn from a complex Gaussian distribution with mean $\bar{\boldsymbol{a}}_s$ (which corresponds to the "nominal" presumed steering vector) and a known covariance matrix $\boldsymbol{C}_a$:

$$\boldsymbol{a}_s \sim \mathcal{CN}(\bar{\boldsymbol{a}}_s, \boldsymbol{C}_a). \tag{2}$$

  The covariance matrix $\boldsymbol{C}_a$ captures the uncertainties in the steering vector. In the satellite communications application, models for the deformations of the reflector are available. These models, along with extensive simulations, allow the simulation of a large number of possible deformations, and therefore, we can obtain (through electromagnetic simulation tools) the associated spatial signatures, from which an accurate estimate of $\boldsymbol{C}_a$ can be made available. We would like to stress that (2) provides a general framework to work with and that it enables us to treat a large number of problems.
- The noise component is a zero-mean, complex-valued Gaussian process with known covariance matrix $\boldsymbol{C}$, i.e.,

$$\boldsymbol{n}_t \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C}).$$

  In the satellite communications application, although frequency reuse is present, adjacent spots will not share the same frequency. Hence, for the spot of interest, interferences can only impinge from spots that are far apart; in such a case, the feeds of the FAFR corresponding to the spot of interest are likely not to receive much energy from these interferences compared with the thermal noise. Even so, one knows where this energy comes from, and therefore, auxiliary feeds can be used to obtain accurate estimates of the interferences covariance matrix. Accordingly, in some operating modes where no communications originate from the spot of interest, one may acquire the noise covariance matrix. The main source of noise is, in fact, the thermal white noise whose power can be accurately measured. However, in order to keep the analysis as general as possible, we consider a general matrix $\boldsymbol{C}$.

Our goal here is to estimate the emitted waveform vector $\boldsymbol{s} = [s(1) \ \cdots \ s(N)]^T$ from N snapshots $\{\boldsymbol{x}_t\}_{t=1}^N$. Toward this end, two approaches are proposed corresponding to two different assumptions for $\boldsymbol{s}$. First, $\boldsymbol{s}$ is considered as a deterministic vector, and a maximum likelihood (ML) approach is presented. The second approach consists of assuming that $\boldsymbol{s}$ is a random vector with some *a priori* probability density function. In such a case, we successively examine a minimum mean-square error (MMSE) and maximum *a posteriori* (MAP) approaches.

Before examining new approaches, it is worthwhile to make the following observation. The most popular approach to solving this problem is to use the minimum variance distortionless beamformer (MVDR). The latter amounts to maximizing the output signal-to-interference plus noise ratio and is given, up to a scaling factor, by

$$\begin{aligned} \boldsymbol{w}_{\mathrm{MVDR}} &\propto \mathcal{P}\left\{\boldsymbol{C}^{-1}\mathcal{E}\left\{|s_t|^2\boldsymbol{a}_s\boldsymbol{a}_s^H\right\}\right\} \\ &\propto \mathcal{P}\left\{\boldsymbol{C}^{-1}\left(\bar{\boldsymbol{a}}_s\bar{\boldsymbol{a}}_s^H+\boldsymbol{C}_a\right)\right\} \end{aligned} \tag{3}$$

where $\mathcal{P}\{\cdot\}$ stands for the principal eigenvector of the matrix between braces. Note that the MVDR beamformer can be computed as $\bar{\boldsymbol{a}}_s, \boldsymbol{C}$ and $\boldsymbol{C}_a$ are known.

III. Maximum Likelihood Estimation

In this section, we derive the maximum likelihood estimator (MLE) of $\boldsymbol{s}$. It will be shown that the latter corresponds to a beamforming-like operation applied to the snapshots. However, the weight vector depends on $\boldsymbol{s}$, which is unknown; in other words, $\boldsymbol{s}$ satisfies an implicit equation of the type $\boldsymbol{s} = \boldsymbol{X}^T\boldsymbol{w}^*(\boldsymbol{s})$. In order to solve this problem, an iterative procedure is proposed. As a first step toward deriving the ML estimator, the probability density function of the observations $\boldsymbol{X} = [\boldsymbol{x}_1 \ \cdots \ \boldsymbol{x}_N]$ is required. For the sake of clarity, these derivations have been deferred to Appendix A, and we refer to this Appendix for details that would be omitted in the sequel. First, note that

$$p(\boldsymbol{X};\boldsymbol{s}) = \int p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s$$

where $p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})$ is the conditional pdf, given $\boldsymbol{a}_s$, $p(\boldsymbol{a}_s)$ is the pdf of the steering vector of interest and where we used a semicolon to stress that the pdf depends on $\boldsymbol{s}$. Using the previous expression along with (27), it follows that

$$\begin{aligned} &\frac{\partial p(\boldsymbol{X};\boldsymbol{s})}{\partial \boldsymbol{s}} \\ &= \int \frac{\partial p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})}{\partial \boldsymbol{s}}p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s \\ &= \int p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})\left\{\boldsymbol{X}^T\boldsymbol{C}^{-1}\boldsymbol{a}_s^* - \boldsymbol{s}\left(\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s\right)\right\}p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s \\ &= -\boldsymbol{s}\left[\int\left(\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s\right)p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s\right] \\ &\quad + \boldsymbol{X}^T\boldsymbol{C}^{-1}\int \boldsymbol{a}_s^*p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s. \end{aligned} \tag{4}$$

Therefore, the MLE of $\boldsymbol{s}$ obeys the following equation:

$$\boldsymbol{s} = \frac{\boldsymbol{X}^T\boldsymbol{C}^{-1}\int \boldsymbol{a}_s^*p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s}{\int\left(\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s\right)p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s} \tag{5}$$

which corresponds to beamformer-like data processing with weight vector given by

$$\begin{aligned} \boldsymbol{w} &= \frac{\boldsymbol{C}^{-1}\int \boldsymbol{a}_s p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s}{\int\left(\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s\right)p(\boldsymbol{X}\,|\,\boldsymbol{a}_s;\boldsymbol{s})p(\boldsymbol{a}_s)\,d\boldsymbol{a}_s} \\ &= \frac{\boldsymbol{C}^{-1}\int \boldsymbol{a}_s p(\boldsymbol{a}_s\,|\,\boldsymbol{X};\boldsymbol{s})\,d\boldsymbol{a}_s}{\int\left(\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s\right)p(\boldsymbol{a}_s\,|\,\boldsymbol{X};\boldsymbol{s})\,d\boldsymbol{a}_s} \\ &= \frac{\boldsymbol{C}^{-1}\boldsymbol{a}_{\mathrm{post}}}{\int\left(\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s\right)p(\boldsymbol{a}_s\,|\,\boldsymbol{X};\boldsymbol{s})\,d\boldsymbol{a}_s} \end{aligned} \tag{6}$$

where $\boldsymbol{a}_{\mathrm{post}} \stackrel{\triangle}{=} \mathcal{E}\{\boldsymbol{a}_s|\boldsymbol{X};\boldsymbol{s}\}$ is the *a posteriori* mean. It should be observed that the previous equation is meaningful; the ML beamformer is, up to a scaling factor, given by $\boldsymbol{C}^{-1}$ times the MMSE estimator of $\boldsymbol{a}_s$. It can also be noticed that when $\boldsymbol{C}_a \to \boldsymbol{0}$, this beamformer converges (up to a scaling factor) to $\boldsymbol{C}^{-1}\bar{\boldsymbol{a}}_s$, which corresponds to the MVDR beamformer for $\boldsymbol{C}_a = \boldsymbol{0}$. Next, since $\boldsymbol{a}_s$ and $\boldsymbol{X}$ are jointly Gaussian, it follows that [27]

$$\boldsymbol{a}_{\mathrm{post}} = \left(\boldsymbol{s}^H\boldsymbol{s}\boldsymbol{C}^{-1}+\boldsymbol{C}_a^{-1}\right)^{-1}\left(\boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s+\boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{s}^*\right). \tag{7}$$

Accordingly, the *a posteriori* covariance matrix is given by

$$\boldsymbol{C}_{\boldsymbol{a}_s|\boldsymbol{X}} = \left(\boldsymbol{s}^H\boldsymbol{s}\boldsymbol{C}^{-1}+\boldsymbol{C}_a^{-1}\right)^{-1} \tag{8}$$

which yields the following expression for $\boldsymbol{w}$:

$$\boldsymbol{w} = \frac{\boldsymbol{C}^{-1}\boldsymbol{a}_{\mathrm{post}}}{\boldsymbol{a}_{\mathrm{post}}^H\boldsymbol{C}^{-1}\boldsymbol{a}_{\mathrm{post}}+\mathrm{Tr}\{\boldsymbol{C}_{\boldsymbol{a}_s|\boldsymbol{X}}\boldsymbol{C}^{-1}\}}. \tag{9}$$

Unfortunately, $\boldsymbol{w}$ depends on $\boldsymbol{s}$, which is unknown, and thus, this beamformer cannot be computed directly from (9). In other words, the ML problem amounts to an implicit equation of the type $\boldsymbol{s} = \boldsymbol{X}^T\boldsymbol{w}^*(\boldsymbol{s})$. However, an iterative algorithm can be employed in order to solve (5). Toward this end, note that $\boldsymbol{s}^* = \boldsymbol{X}^H\boldsymbol{w}$ so that

$$\begin{aligned} &\boldsymbol{s}^H\boldsymbol{s} = N\boldsymbol{w}^H\hat{\boldsymbol{R}}\boldsymbol{w}; \quad \boldsymbol{X}\boldsymbol{s}^* = N\hat{\boldsymbol{R}}\boldsymbol{w} \\ &\boldsymbol{a}_{\mathrm{post}} = \left(N\boldsymbol{w}^H\hat{\boldsymbol{R}}\boldsymbol{w}\boldsymbol{C}^{-1}+\boldsymbol{C}_a^{-1}\right)^{-1}\left(\boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s+N\boldsymbol{C}^{-1}\hat{\boldsymbol{R}}\boldsymbol{w}\right) \end{aligned}$$

where

$$\hat{\boldsymbol{R}} = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^H \tag{10}$$

denotes the sample covariance matrix. The previous relations suggest the iterative scheme of Table I to estimate $\boldsymbol{w}$.

This scheme allows the recovery of the solution of the ML problem. The algorithm requires approximately $6m^2$ multiplications/additions plus a matrix inversion at each iteration. The algorithm is stopped whenever $\|\boldsymbol{w}^{(n+1)} - \boldsymbol{w}^{(n)}\| \leq \delta$, where $\delta = 0.001/\sqrt{m}$ in the simulations of Section V. Note that in the case of small uncertainties, the ML solution should not be far from the MVDR beamformer, and hence, initializing the iterative scheme with the MVDR solution already provides a good initial estimate, which prevents convergence problems. Indeed, in the numerical results shown in Section V, we did not

TABLE I
ITERATIVE SCHEME FOR COMPUTING THE ML BEAMFORMER

Initialisation: $\boldsymbol{w}^{(1)} = \boldsymbol{w}_{\text{MVDR}}$; $P^{(1)} = \boldsymbol{w}^{(1)H}\widehat{\boldsymbol{R}}\boldsymbol{w}^{(1)}$
Recursion: for $n = 1, \cdots,$
$\boldsymbol{a}_{\text{post}}^{(n+1)} = \left(NP^{(n)}\boldsymbol{C}^{-1} + \boldsymbol{C}_a^{-1}\right)^{-1}\left(\boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s + N\boldsymbol{C}^{-1}\widehat{\boldsymbol{R}}\boldsymbol{w}^{(n)}\right)$
$\boldsymbol{w}^{(n+1)} = \dfrac{\boldsymbol{C}^{-1}\boldsymbol{a}_{\text{post}}^{(n+1)}}{\boldsymbol{a}_{\text{post}}^{(n+1)H}\boldsymbol{C}^{-1}\boldsymbol{a}_{\text{post}}^{(n+1)} + \text{Tr}\left\{\left(NP^{(n)}\boldsymbol{C}^{-1}+\boldsymbol{C}_a^{-1}\right)^{-1}\boldsymbol{C}^{-1}\right\}}$
$P^{(n+1)} = \boldsymbol{w}^{(n+1)H}\widehat{\boldsymbol{R}}\boldsymbol{w}^{(n+1)}$

encounter any convergence problem [we observed that convergence is achieved within a few (two to ten) iterations].

IV. Bayesian Approaches

In this section, we consider a Bayesian approach in which $\boldsymbol{s}$ is considered to be a random vector we would like to estimate. For the sake of convenience, we assume that $\boldsymbol{s}$ is complex Gaussian, with zero-mean and a covariance matrix equal to $\sigma_s^2\boldsymbol{I}_N$.

A. Approximate MMSE Estimation

First, we investigate a MMSE approach. The MMSE estimator of s_t can be written as [27]

$$\begin{aligned}\hat{s}_t &= \mathcal{E}\{s_t \,|\, \boldsymbol{X}\} = \int s_t p(s_t \,|\, \boldsymbol{X})\, ds_t \\ &= \int s_t \left[\int p(s_t \,|\, \boldsymbol{X}, \boldsymbol{a}_s) p(\boldsymbol{a}_s \,|\, \boldsymbol{X})\, d\boldsymbol{a}_s\right] ds_t \\ &= \int p(\boldsymbol{a}_s \,|\, \boldsymbol{X}) \left[\int s_t p(s_t \,|\, \boldsymbol{X}, \boldsymbol{a}_s)\, ds_t\right] d\boldsymbol{a}_s \\ &= \sigma_s^2 \int \left(\boldsymbol{a}_s^H \boldsymbol{R}^{-1}\boldsymbol{x}_t\right) p(\boldsymbol{a}_s \,|\, \boldsymbol{X})\, d\boldsymbol{a}_s \\ &= \boldsymbol{w}_{\text{MMSE}}^H \boldsymbol{x}_t \end{aligned} \tag{11}$$

where $\boldsymbol{R} = \sigma_s^2\boldsymbol{a}_s\boldsymbol{a}_s^H + \boldsymbol{C}$ is the covariance matrix for a given $\boldsymbol{a}_s$. In the previous equation, we used the fact that $\int s_t p(s_t \,|\, \boldsymbol{X}, \boldsymbol{a}_s)\, ds_t$ is the MMSE estimate of s_t for a given $\boldsymbol{a}_s$. When $\boldsymbol{a}_s$ is known, $\boldsymbol{x}_t$ and s_t are jointly Gaussian, and therefore, the MMSE estimate of s_t becomes $\boldsymbol{a}_s^H\boldsymbol{R}^{-1}\boldsymbol{x}_t$. The MMSE estimator is a beamformer whose weight vector has the following expression:

$$\begin{aligned}\boldsymbol{w}_{\text{MMSE}} &= \sigma_s^2 \int p(\boldsymbol{a}_s|\boldsymbol{X})\boldsymbol{R}^{-1}\boldsymbol{a}_s\, d\boldsymbol{a}_s \\ &= \sigma_s^2 \int \frac{\boldsymbol{C}^{-1}\boldsymbol{a}_s}{1+\sigma_s^2\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s} p(\boldsymbol{a}_s \,|\, \boldsymbol{X})\, d\boldsymbol{a}_s \\ &= \frac{\sigma_s^2}{p(\boldsymbol{X})} \int \frac{\boldsymbol{C}^{-1}\boldsymbol{a}_s}{1+\sigma_s^2\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s} p(\boldsymbol{X} \,|\, \boldsymbol{a}_s) p(\boldsymbol{a}_s)\, d\boldsymbol{a}_s. \end{aligned} \tag{12}$$

Unfortunately, it appears that no exact expression is available for the above integral, and one has to resort to some approximations; see, e.g., [20] for a similar discussion. Prior to that, a few observations are in order. When the signal-to-noise ratio (SNR) is large, or when N is large or when there are high uncertainties, the *a posteriori* probability $p(\boldsymbol{a}_s \,|\, \boldsymbol{X})$ will tend to be highly concentrated around the *a posteriori* mean of the steering vector. Under such conditions, the MMSE weight vector will approximately be proportional to $\boldsymbol{C}^{-1}\mathcal{E}\{\boldsymbol{a}_s \,|\, \boldsymbol{X}\}$. In contrast, when the SNR is low and N or $\boldsymbol{C}_a$ is small, the *a posteriori* density will be nearly equal to the *a priori* pdf, and therefore

$$\boldsymbol{w}_{\text{MMSE}} \simeq \int \frac{\sigma_s^2\boldsymbol{C}^{-1}\boldsymbol{a}_s}{1+\sigma_s^2\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s} p(\boldsymbol{a}_s)\, d\boldsymbol{a}_s. \tag{13}$$

The weight vector in (13) has also a nice interpretation. It corresponds to averaging, over the pdf of $\boldsymbol{a}_s$, the optimal weight vector for a given $\boldsymbol{a}_s$. Therefore, it corresponds to a sensible approach. In either case (12) or (13), one needs to compute

$$\int \frac{\boldsymbol{a}_s}{1+\sigma_s^2\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s} \tilde{p}(\boldsymbol{a}_s)\, d\boldsymbol{a}_s \tag{14}$$

where $\tilde{p}(\boldsymbol{a}_s)$ is either the *a posteriori* or the *a priori* density of $\boldsymbol{a}_s$. In Appendix B, we show that this integral can be accurately approximated by

$$\int \frac{\boldsymbol{a}_s\tilde{p}(\boldsymbol{a}_s)\, d\boldsymbol{a}_s}{1+\sigma_s^2\boldsymbol{a}_s^H\boldsymbol{C}^{-1}\boldsymbol{a}_s} \simeq \alpha\tilde{\mathcal{E}}\{\boldsymbol{a}_s\} + \beta\tilde{\boldsymbol{C}}_{\boldsymbol{a}_s}\boldsymbol{C}^{-1}\tilde{\mathcal{E}}\{\boldsymbol{a}_s\} \tag{15}$$

where $\tilde{\mathcal{E}}\{\boldsymbol{a}_s\}$ is the mean of $\boldsymbol{a}_s$ (associated with $\tilde{p}$), and $\tilde{\boldsymbol{C}}_{\boldsymbol{a}_s}$ is the corresponding covariance matrix, where α and β depend on $\sigma_s^2, \boldsymbol{C}^{-1}, \tilde{\mathcal{E}}\{\boldsymbol{a}_s\}$ and $\tilde{\boldsymbol{C}}_{\boldsymbol{a}_s}$. It appears rather difficult to derive the mean and the covariance matrix of $\boldsymbol{a}_s|\boldsymbol{X}$. Indeed, we get from (7)

$$\begin{aligned}\mathcal{E}\{\boldsymbol{a}_s \,|\, \boldsymbol{X}\} = \mathcal{E}_{\boldsymbol{s}}\Big\{&\left(\boldsymbol{s}^H\boldsymbol{s}\boldsymbol{C}^{-1} + \boldsymbol{C}_a^{-1}\right)^{-1} \\ &\times \left(\boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s + \boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{s}^*\right)\Big\}\end{aligned} \tag{16}$$

and therefore, $\mathcal{E}\{\boldsymbol{a}_s \,|\, \boldsymbol{X}\}$ is not obvious to obtain. In addition, as mentioned above, a more practically interesting situation is when the SNR is low or N is small; under these conditions, the *a posteriori* will approximately be the *a priori* density. Therefore, we propose to compute an approximate MMSE beamformer from (13). Using (15), it leads to

$$\begin{aligned}\boldsymbol{w} &= \sigma_s^2\boldsymbol{C}^{-1}[\alpha\bar{\boldsymbol{a}}_s + \beta\boldsymbol{C}_a\boldsymbol{C}^{-1}\bar{\boldsymbol{a}}_s] \\ &= \sigma_s^2(\alpha\boldsymbol{I}_m + \beta\boldsymbol{C}^{-1}\boldsymbol{C}_a)\boldsymbol{C}^{-1}\bar{\boldsymbol{a}}_s.\end{aligned} \tag{17}$$

Note that it corresponds to a weighted sum of the optimal beamformer in the absence of uncertainty and a term that depends on the level of uncertainty. Finally, observe that in practice, σ_s^2 may not be known, and therefore, (17) cannot be implemented directly. In order to remedy this problem, a simple solution is to estimate the source power. The most intuitive and simple way in practice is to set

$$\hat{\sigma_s^2} = \bar{\boldsymbol{w}}^H\hat{\boldsymbol{R}}\bar{\boldsymbol{w}} = \frac{\bar{\boldsymbol{a}}_s^H\boldsymbol{C}^{-1}\hat{\boldsymbol{R}}\boldsymbol{C}^{-1}\bar{\boldsymbol{a}}_s}{\left(\bar{\boldsymbol{a}}_s^H\boldsymbol{C}^{-1}\bar{\boldsymbol{a}}_s\right)^2} \tag{18}$$

which corresponds to the optimal source power estimate in the absence of uncertainties. In case of small uncertainties, this estimate should remain quite accurate. In summary, the approximate MMSE weight vector is given by (17), where σ_s^2 is replaced by its estimate in (18).

B. Maximum a Posteriori Estimation

Since the MMSE cannot be implemented directly and only an approximation of it is available, we turn to the derivation of the MAP estimator. The latter is obtained as the solution to the following maximization problem:

$$\hat{\boldsymbol{s}} = \arg\max_{\boldsymbol{s}} p(\boldsymbol{s}|\boldsymbol{X}) = \arg\max_{\boldsymbol{s}} p(\boldsymbol{X} \,|\, \boldsymbol{s}) p(\boldsymbol{s}). \tag{19}$$

Observing that

$$\begin{aligned} p(\boldsymbol{X} \,|\, \boldsymbol{s}) &= \int p(\boldsymbol{X} \,|\, \boldsymbol{a}_s, \boldsymbol{s}) p(\boldsymbol{a}_s)\, d\boldsymbol{a}_s \\ \frac{\partial p(\boldsymbol{X} \,|\, \boldsymbol{a}_s, \boldsymbol{s})}{\partial \boldsymbol{s}} &= p(\boldsymbol{X} \,|\, \boldsymbol{a}_s, \boldsymbol{s}) \left\{ \boldsymbol{X}^T \boldsymbol{C}^{-1} \boldsymbol{a}_s^* - \boldsymbol{s} \left(\boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s \right) \right\} \\ \frac{\partial p(\boldsymbol{s})}{\partial \boldsymbol{s}} &= -\frac{\boldsymbol{s}}{\sigma_s^2} p(\boldsymbol{s}) \end{aligned} \tag{20}$$

it follows that

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{s}} [p(\boldsymbol{X}|\boldsymbol{s}) p(\boldsymbol{s})] \\ &\quad = \frac{\partial p(\boldsymbol{X} \,|\, \boldsymbol{s})}{\partial \boldsymbol{s}} p(\boldsymbol{s}) + p(\boldsymbol{X} \,|\, \boldsymbol{s}) \frac{\partial p(\boldsymbol{s})}{\partial \boldsymbol{s}} \\ &\quad = p(\boldsymbol{s}) \left\{ \frac{\partial p(\boldsymbol{X} \,|\, \boldsymbol{s})}{\partial \boldsymbol{s}} - \frac{\boldsymbol{s}}{\sigma_s^2} p(\boldsymbol{X} \,|\, \boldsymbol{s}) \right\} \\ &\quad = p(\boldsymbol{s}) \left\{ \int \left[\boldsymbol{X}^T \boldsymbol{C}^{-1} \boldsymbol{a}_s^* - \boldsymbol{s} \left(\boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s \right) - \frac{\boldsymbol{s}}{\sigma_s^2} \right] \right. \\ &\qquad \left. \times\, p(\boldsymbol{X} \,|\, \boldsymbol{a}_s, \boldsymbol{s}) p(\boldsymbol{a}_s)\, d\boldsymbol{a}_s \right\}. \end{aligned} \tag{21}$$

Consequently, using arguments similar to those leading to the ML estimator, the MAP estimator is also in a beamforming-like form, where the weight vector is given by

$$\begin{aligned} \boldsymbol{w} &= \frac{\boldsymbol{C}^{-1} \int \boldsymbol{a}_s p(\boldsymbol{X}|\boldsymbol{a}_s, \boldsymbol{s}) p(\boldsymbol{a}_s)\, d\boldsymbol{a}_s}{\int \left[\boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s + \sigma_s^{-2} \right] p(\boldsymbol{X} \,|\, \boldsymbol{a}_s, \boldsymbol{s}) p(\boldsymbol{a}_s)\, d\boldsymbol{a}_s} \\ &= \frac{\boldsymbol{C}^{-1} \mathcal{E}\{\boldsymbol{a}_s | \boldsymbol{X}, \boldsymbol{s}\}}{\int \left[\boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s + \sigma_s^{-2} \right] p(\boldsymbol{a}_s \,|\, \boldsymbol{X}, \boldsymbol{s})\, d\boldsymbol{a}_s} \\ &= \frac{\boldsymbol{C}^{-1} \mathcal{E}\{\boldsymbol{a}_s | \boldsymbol{X}, \boldsymbol{s}\}}{\mathrm{Tr}\left\{ \boldsymbol{R}_{\boldsymbol{a}_s \,|\, \boldsymbol{X}, \boldsymbol{s}} \boldsymbol{C}^{-1} \right\} + \sigma_s^{-2}} \end{aligned} \tag{22}$$

where $\boldsymbol{R}_{\boldsymbol{a}_s \,|\, \boldsymbol{X}, \boldsymbol{s}}$ is the *a posteriori* correlation matrix given $\boldsymbol{X}$ and $\boldsymbol{s}$. The previous equation is very similar to that satisfied by the ML weight vector [see (6)] except for the term σ_s^{-2} in the denominator. This additional term may induce some differences between the ML and the MAP estimator, especially at low SNR. In order to obtain the MAP estimate, an iterative procedure, similar to that proposed in Section III—see Table I—for the ML estimator can be employed with the slight modification mentioned above regarding σ_s^{-2}. However, the latter quantity is unknown. Similarly to what was done for the MMSE estimator, we propose to replace, at each stage of the iterative scheme, σ_s^2 by an estimate, namely, $\hat{\sigma_s^2} = P^{(n)}$. Therefore, the scheme of Table I can be used with a single modification affecting the denominator in the computation of $\boldsymbol{w}^{(n+1)}$.

C. Joint MAP Estimation of $\boldsymbol{a}_s$ and $\boldsymbol{s}$

In the previous subsection, MAP estimation of $\boldsymbol{s}$ only was proposed. However, since both $\boldsymbol{a}_s$ and $\boldsymbol{s}$ are random (and unknown), a possible approach is to estimate them jointly. This is achieved by maximizing $p(\boldsymbol{a}_s, \boldsymbol{s} \,|\, \boldsymbol{X})$ or, equivalently, $p(\boldsymbol{X}|\boldsymbol{a}_s, \boldsymbol{s}) p(\boldsymbol{a}_s, \boldsymbol{s}) = p(\boldsymbol{X} \,|\, \boldsymbol{a}_s, \boldsymbol{s}) p(\boldsymbol{a}_s) p(\boldsymbol{s})$. Given the assumptions made, this amounts to minimizing (see Appendix A for details)

$$\begin{aligned} \Lambda = &-\boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{X} \boldsymbol{s}^* - \boldsymbol{s}^T \boldsymbol{X}^H \boldsymbol{C}^{-1} \boldsymbol{a}_s + \left(\boldsymbol{s}^H \boldsymbol{s} \right) \left(\boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s \right) \\ &+ (\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)^H \boldsymbol{C}_a^{-1} (\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s) + \sigma_s^{-2} \boldsymbol{s}^H \boldsymbol{s}. \end{aligned} \tag{23}$$

Differentiating Λ w.r.t. $\boldsymbol{s}$ yields

$$\boldsymbol{s} = \boldsymbol{X}^T \boldsymbol{w}^*; \quad \boldsymbol{w} = \frac{\sigma_s^2 \boldsymbol{C}^{-1} \boldsymbol{a}_s}{1 + \sigma_s^2 \boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s}. \tag{24}$$

Inserting this value into Λ, the MAP estimate of $\boldsymbol{a}_s$ is obtained by minimizing

$$\begin{aligned} \tilde{\Lambda} = &-\frac{\sigma_s^2 \boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{X} \boldsymbol{X}^H \boldsymbol{C}^{-1} \boldsymbol{a}_s}{1 + \sigma_s^2 \boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s} \\ &+ (\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)^H \boldsymbol{C}_a^{-1} (\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s). \end{aligned} \tag{25}$$

This corresponds to an m-dimensional minimization problem. In order to solve it efficiently, first note that the gradient is easily computed as

$$\begin{aligned} \frac{\partial \tilde{\Lambda}}{\partial \boldsymbol{a}_s} = &\, \boldsymbol{C}_a^{-1} (\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s) - \sigma_s^2 \frac{\boldsymbol{C}^{-1} \boldsymbol{X} \boldsymbol{X}^H \boldsymbol{C}^{-1} \boldsymbol{a}_s}{1 + \sigma_s^2 \boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s} \\ &- \sigma_s^2 \frac{\left(\boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{X} \boldsymbol{X}^H \boldsymbol{C}^{-1} \boldsymbol{a}_s \right) \boldsymbol{C}^{-1} \boldsymbol{a}_s}{(1 + \sigma_s^2 \boldsymbol{a}_s^H \boldsymbol{C}^{-1} \boldsymbol{a}_s)^2}. \end{aligned}$$

Therefore, computationally efficient gradient-based methods can be used to minimize this function. However, a good initial estimate is necessary to avoid convergence to local minima and to avoid running too many iterations. An initial estimate can be found by observing that $\bar{\boldsymbol{a}}_s$ cancels the second term of $\tilde{\Lambda}$. Additionally, for small $\boldsymbol{C}_a$, the minimizer of $\tilde{\Lambda}$ should be close to $\bar{\boldsymbol{a}}_s$. When the uncertainty increases, however, emphasis should be placed on minimizing the first term of $\tilde{\Lambda}$. This is approximately equivalent to selecting $\boldsymbol{a}_s$ as the principal eigenvector of $\boldsymbol{C}^{-1}\hat{\boldsymbol{R}}$. Combining these two observations, we choose as an initial guess for $\boldsymbol{a}_s$ the projection of $\bar{\boldsymbol{a}}_s$ onto the principal eigenvector of $\boldsymbol{C}^{-1}\hat{\boldsymbol{R}}$. In the numerical results to be presented next, this choice turns out to be effective in that we did not encounter numerical problems with minimizing $\tilde{\Lambda}$.

V. Numerical Examples

In this section, we provide numerical illustrations of the performances of the four estimators proposed, namely, the ML beamformer (9), implemented as in Table I, the approximate MMSE (17), and the MAP estimators. For comparison purposes, we also display the performance of the MVDR beamformer (3). We consider a FAFR where $m = 7$ feeds are used to form a spot. The matrix $\boldsymbol{C}_a$ was computed through electromagnetic simulations, taking into account various kinds of deformations that can affect the paraboloid. In all

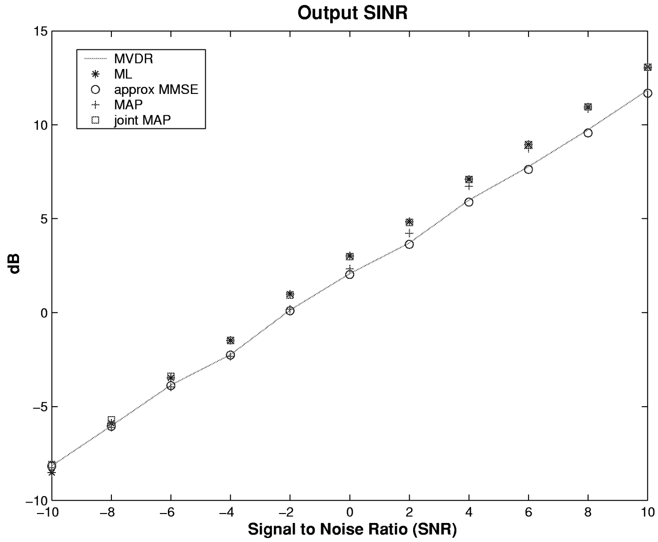

Fig. 1. Output SINR versus SNR. $N = 50$ and UR $= 0$ dB.

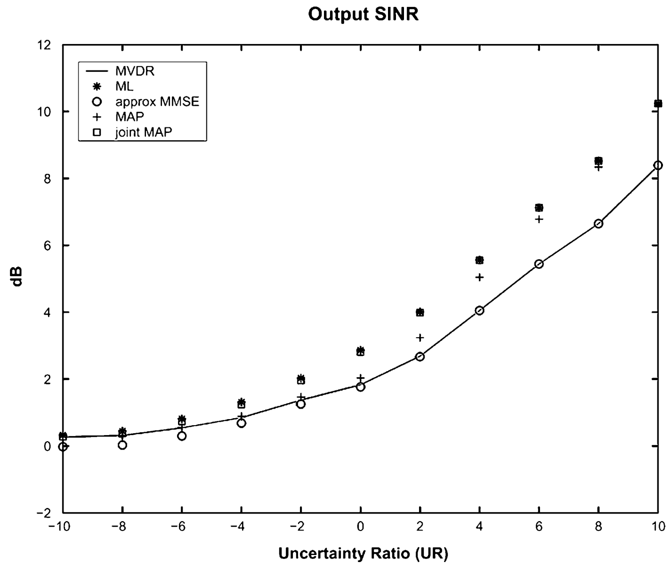

Fig. 2. Output SINR versus UR. $N = 50$ and SNR $= 0$ dB.

simulations, the noise covariance matrix will be $\boldsymbol{C} = \sigma_n^2 \boldsymbol{I}_m$. The source signal s_t is generated as a random Gaussian process with power σ_s^2. The number of snapshots is set to $N = 50$. We define the uncertainty ratio and the SNR, respectively, as

$$\mathrm{UR} = 10 \log_{10} \left(\frac{\mathrm{Tr}\{\boldsymbol{C}_a\}}{\bar{\boldsymbol{a}}_s^H \bar{\boldsymbol{a}}_s} \right)$$
$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\sigma_s^2 \bar{\boldsymbol{a}}_s^H \bar{\boldsymbol{a}}_s}{\sigma_n^2} \right).$$

All estimators will be compared in terms of the output signal-to-interference-plus-noise ratio (SINR). For each figure, $N_r = 300$ Monte Carlo simulations were run with a different $\boldsymbol{a}_s$ drawn from (2), and a weight vector was computed using the methods presented above. For a generic weight vector $\boldsymbol{w}$, the SINR is averaged and computed as

$$\mathrm{SINR}(\boldsymbol{w}) = \frac{1}{N_r} \sum_{n=1}^{N_r} \frac{\sigma_s^2 \left| \boldsymbol{w}^H \boldsymbol{a}_s(n) \right|^2}{\boldsymbol{w}^H \boldsymbol{C} \boldsymbol{w}}$$

and is plotted in decibels (dB) in the figures. We examine the performances of the proposed estimators versus the SNR and the uncertainty ratio; see Figs. 1 and 2, respectively.

The following comments can be made from inspection of these figures. The ML, MAP, and joint MAP beamformer are seen to produce the most accurate estimates in nearly all scenarios, with the ML and joint MAP beamformers performing slightly better. These two beamformers outperform the MVDR beamformer, especially for high uncertainty ratio or high SNR. For instance, at SNR $= 0$ dB, the improvement is around 1 dB for UR $= 0$ dB and 2 dB for UR $= 10$ dB. Similarly, for UR $= 0$ dB, the improvement is around 1 dB for SNR superior to -2 dB. In case of small UR, the MVDR beamformer performs similarly as these three methods. Observe that the ML and joint MAP estimators provide very good output SINR, even in short data samples or high UR. The approximate MMSE estimator provides an output SINR inferior to that of the ML and MAP estimators. This result may be explained by the fact that it approximates the *a posteriori* probability $p(\boldsymbol{a}_s|\boldsymbol{X})$ by the *a priori* pdf $p(\boldsymbol{a}_s)$. Therefore, it does not make use of the information about $\boldsymbol{a}_s$ brought by the data.

VI. Conclusion

In this paper, we considered the problem of estimating a signal waveform when there exist uncertainties about the spatial signature of interest, but at the same time, some *a priori* information about these uncertainties is available. To account for this information, a probability density function was assumed for the random steering vector. This provides a general framework that can accommodate many practical cases of interest. In this preliminary study, we consider the case where the interference plus noise covariance matrix is known. Deterministic (i.e., maximum likelihood) as well as Bayesian approaches were investigated. It was shown that they all entail finding a beamformer. For the ML and MAP estimators, an iterative scheme was proposed, whereas the approximate MMSE beamformer is obtained in closed form. The joint MAP estimator requires the minimization of an m-D function. The proposed methods were shown to provide accurate estimates, especially the ML and joint MAP beamformers, even in the presence of high uncertainties. A future direction of research consists of deriving "adaptive" versions of these algorithms, considering a possibly unknown noise covariance matrix $\boldsymbol{C}$.

Appendix A
Derivation of the Probability Density Function of the Observations

In this Appendix, we provide an expression for the pdf of the observations that, in turn, will be used to derive the maximum likelihood estimator. Toward this end, we can write

$$p(\boldsymbol{X}; \boldsymbol{s}) = \int p(\boldsymbol{X} \mid \boldsymbol{a}_s; \boldsymbol{s}) p(\boldsymbol{a}_s)\, d\boldsymbol{a}_s \tag{26}$$

where $p(\boldsymbol{X} \mid \boldsymbol{a}_s; \boldsymbol{s})$ is the conditional pdf for a given $\boldsymbol{a}_s$, whereas $p(\boldsymbol{a}_s)$ denotes the *a priori* pdf of the steering vector of interest.

The semicolon indicates that the pdf depends on $\boldsymbol{s}$. From the assumptions made, the conditional pdf is given by

$$p(\boldsymbol{X} \mid \boldsymbol{a}_s; \boldsymbol{s}) = \frac{e^{-\sum_{t=1}^{N}(\boldsymbol{x}_t - \boldsymbol{a}_s s_t)^H \boldsymbol{C}^{-1}(\boldsymbol{x}_t - \boldsymbol{a}_s s_t)}}{\pi^{mN} |\boldsymbol{C}|^N} \tag{27}$$

where $|\boldsymbol{C}|$ stands for the determinant of matrix $\boldsymbol{C}$. By some straightforward algebra, one can show that

$$\begin{aligned}\sum_{t=1}^{N}&(\boldsymbol{x}_t - \boldsymbol{a}_s s_t)^H \boldsymbol{C}^{-1}(\boldsymbol{x}_t - \boldsymbol{a}_s s_t) \\ &= (\boldsymbol{s}^H\boldsymbol{s})\left(\boldsymbol{a}_s - \frac{\boldsymbol{X}\boldsymbol{s}^*}{\boldsymbol{s}^H\boldsymbol{s}}\right)^H \boldsymbol{C}^{-1}\left(\boldsymbol{a}_s - \frac{\boldsymbol{X}\boldsymbol{s}^*}{\boldsymbol{s}^H\boldsymbol{s}}\right) + \xi \end{aligned} \tag{28}$$

with $\xi = -(\boldsymbol{s}^T\boldsymbol{X}^H\boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{s}^*/\boldsymbol{s}^H\boldsymbol{s}) + \mathrm{Tr}\{\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{C}^{-1}\}$ so that

$$\begin{aligned} p(\boldsymbol{X}; \boldsymbol{s}) &= \int \frac{1}{\pi^m |\boldsymbol{C}_a|} p(\boldsymbol{X} \mid \boldsymbol{a}_s; \boldsymbol{s}) \\ &\quad \times e^{-(\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)^H \boldsymbol{C}_a^{-1}(\boldsymbol{a}_s - \bar{a}_s)}\, d\boldsymbol{a}_s \\ &= \frac{e^{-\xi}}{\pi^{m(N+1)}|\boldsymbol{C}|^N|\boldsymbol{C}_a|} \\ &\quad \times \int e^{-(\boldsymbol{s}^H\boldsymbol{s})\left(\boldsymbol{a}_s - \frac{\boldsymbol{X}\boldsymbol{s}^*}{\boldsymbol{s}^H\boldsymbol{s}}\right)^H \boldsymbol{C}^{-1}\left(\boldsymbol{a}_s - \frac{\boldsymbol{X}\boldsymbol{s}^*}{\boldsymbol{s}^H\boldsymbol{s}}\right)} \\ &\quad \times e^{-(\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)^H \boldsymbol{C}_a^{-1}(\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)}\, d\boldsymbol{a}_s. \end{aligned} \tag{29}$$

Next, for any vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in $\mathbb{C}^m$ and any two positive definite covariance matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$, the following relation is readily shown to hold:

$$\begin{aligned}(\boldsymbol{a} - \boldsymbol{a}_1)^H\boldsymbol{C}_1^{-1}(\boldsymbol{a} - \boldsymbol{a}_1)^H &+ (\boldsymbol{a} - \boldsymbol{a}_2)^H\boldsymbol{C}_2^{-1}(\boldsymbol{a} - \boldsymbol{a}_2) \\ &= (\boldsymbol{a} - \tilde{\boldsymbol{C}}\tilde{\boldsymbol{a}})^H\tilde{\boldsymbol{C}}^{-1}(\boldsymbol{a} - \tilde{\boldsymbol{C}}\tilde{\boldsymbol{a}}) \\ &\quad + \boldsymbol{a}_1^H\boldsymbol{C}_1^{-1}\boldsymbol{a}_1 + \boldsymbol{a}_2^H\boldsymbol{C}_2^{-1}\boldsymbol{a}_2 - \tilde{\boldsymbol{a}}^H\tilde{\boldsymbol{C}}\tilde{\boldsymbol{a}}\end{aligned}$$

with $\tilde{\boldsymbol{C}}^{-1} = \boldsymbol{C}_1^{-1} + \boldsymbol{C}_2^{-1}$ and $\tilde{\boldsymbol{a}} = \boldsymbol{C}_1^{-1}\boldsymbol{a}_1 + \boldsymbol{C}_2^{-1}\boldsymbol{a}_2$. Using this result along with the fact that

$$\int e^{-(\boldsymbol{a}-\boldsymbol{a}_1)^H\boldsymbol{C}_1^{-1}(\boldsymbol{a}-\boldsymbol{a}_1)}\, d\boldsymbol{a} = \pi^m|\boldsymbol{C}_1|$$

yields

$$\begin{aligned} p(\boldsymbol{X}; \boldsymbol{s}) &= \frac{\pi^m e^{-\xi} e^{-(\boldsymbol{s}^H\boldsymbol{s})^{-1}\boldsymbol{s}^T\boldsymbol{X}^H\boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{s}^*} e^{-\bar{\boldsymbol{a}}_s^H\boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s}}{\pi^{m(N+1)}|\boldsymbol{C}|^N|\boldsymbol{C}_a|\left|\boldsymbol{s}^H\boldsymbol{s}\boldsymbol{C}^{-1} + \boldsymbol{C}_a^{-1}\right|} \\ &\quad \times e^{(\boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{s}^* + \boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s)^H(\boldsymbol{s}^H\boldsymbol{s}\boldsymbol{C}^{-1} + \boldsymbol{C}_a^{-1})^{-1}(\boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{s}^* + \boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s)} \\ &= \frac{e^{-\mathrm{Tr}\{\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{C}^{-1}\}} e^{-\bar{\boldsymbol{a}}_s^H\boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s}}{\pi^{mN}|\boldsymbol{C}|^N|\boldsymbol{C}_a|\left|\boldsymbol{s}^H\boldsymbol{s}\boldsymbol{C}^{-1} + \boldsymbol{C}_a^{-1}\right|} \\ &\quad \times e^{(\boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{s}^* + \boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s)^H(\boldsymbol{s}^H\boldsymbol{s}\boldsymbol{C}^{-1} + \boldsymbol{C}_a^{-1})^{-1}(\boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{s}^* + \boldsymbol{C}_a^{-1}\bar{\boldsymbol{a}}_s)} \end{aligned} \tag{30}$$

which completes the derivation of the pdf of the observations.

Remark 1: An alternate derivation for the above expression can be obtained as follows. Let $\boldsymbol{x} = [\boldsymbol{x}^T(1)\ \cdots\ \boldsymbol{x}^T(N)]^T = \mathrm{vec}(\boldsymbol{X})$ be the $Nm \times 1$ vector constructed by stacking the columns of $\boldsymbol{X}$. Since $\boldsymbol{a}_s$ and $\boldsymbol{n}_t$ are independent Gaussian vectors, it follows that $\boldsymbol{x}$ is Gaussian distributed. Additionally, observe that

$$\begin{aligned}\mathcal{E}\{\boldsymbol{x}_t\} &= \mathcal{E}_{\boldsymbol{a}_s}\left\{\mathcal{E}_{|\boldsymbol{a}_s}\{\boldsymbol{x}_t\}\right\} = \mathcal{E}_{\boldsymbol{a}_s}\left\{\mathcal{E}_{|\boldsymbol{a}_s}\{\boldsymbol{a}_s s_t + \boldsymbol{n}_t\}\right\} \\ &= \mathcal{E}_{\boldsymbol{a}_s}\{\boldsymbol{a}_s s_t\} = \bar{\boldsymbol{a}}_s s_t\end{aligned}$$

and similarly

$$\begin{aligned}\mathcal{E}&\{(\boldsymbol{x}_t - \bar{\boldsymbol{a}}_s s_t)(\boldsymbol{x}_v - \bar{\boldsymbol{a}}_s s_v)^H\} \\ &= \mathcal{E}_{\boldsymbol{a}_s}\{\mathcal{E}_{|\boldsymbol{a}_s}\{[(\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)s_t + \boldsymbol{n}_t][(\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)s_v + \boldsymbol{n}_v]^H\}\} \\ &= \mathcal{E}_{\boldsymbol{a}_s}\left\{(\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)(\boldsymbol{a}_s - \bar{\boldsymbol{a}}_s)^H s_t s_v^* + \boldsymbol{C}\delta(t,v)\right\} \\ &= \boldsymbol{C}_a s_t s_v^* + \boldsymbol{C}\delta(t,v).\end{aligned}$$

In the previous equations, $\mathcal{E}_{|\boldsymbol{a}_s}\{\,.\,\}$ stands for the conditional mean given $\boldsymbol{a}_s$, whereas $\mathcal{E}_{\boldsymbol{a}_s}\{\cdot\}$ is the mean with respect to (w.r.t.) the pdf of $\boldsymbol{a}_s$. Hence, $\boldsymbol{x}$ is a Gaussian vector with mean $\boldsymbol{s} \otimes \bar{\boldsymbol{a}}_s$ and covariance matrix $\boldsymbol{s}\boldsymbol{s}^H \otimes \boldsymbol{C}_a + \boldsymbol{I}_N \otimes \boldsymbol{C}$. It follows that its pdf is given by

$$p(\boldsymbol{x}; \boldsymbol{s}) = \frac{e^{-(\boldsymbol{x} - \boldsymbol{s}\otimes\bar{\boldsymbol{a}}_s)^H(\boldsymbol{s}\boldsymbol{s}^H \otimes \boldsymbol{C}_a + \boldsymbol{I}_N \otimes \boldsymbol{C})^{-1}(\boldsymbol{x} - \boldsymbol{s}\otimes\bar{\boldsymbol{a}}_s)}}{\pi^{mN}|\boldsymbol{s}\boldsymbol{s}^H \otimes \boldsymbol{C}_a + \boldsymbol{I}_N \otimes \boldsymbol{C}|}. \tag{31}$$

APPENDIX B
APPROXIMATING THE INTEGRALS IN (12) AND (13)

In this Appendix, we provide an approximation for the MMSE weight vector. Let $\boldsymbol{u}$ be a real-valued random vector with associated pdf $p(\boldsymbol{u})$, and let $\bar{\boldsymbol{u}}$ and $\boldsymbol{C}_{\boldsymbol{u}}$ denote its mean and covariance matrix, respectively. Let us consider the problem of calculating the following integral:

$$I_u = \int f(\boldsymbol{u})p(\boldsymbol{u})\, d\boldsymbol{u} \tag{32}$$

where $f(\cdot)$ is a real-valued scalar function of $\boldsymbol{u}$. Toward this end, we can expand $f(\cdot)$ in a Taylor series around $\bar{\boldsymbol{u}}$ as

$$\begin{aligned} f(\boldsymbol{u}) &= f(\bar{\boldsymbol{u}}) + \left.\frac{\partial f}{\partial \boldsymbol{u}^T}\right|_{\bar{\boldsymbol{u}}}(\boldsymbol{u} - \bar{\boldsymbol{u}}) \\ &\quad + \frac{1}{2}(\boldsymbol{u} - \bar{\boldsymbol{u}})^T.\left.\frac{\partial^2 f}{\partial \boldsymbol{u}\partial \boldsymbol{u}^T}\right|_{\bar{\boldsymbol{u}}}(\boldsymbol{u} - \bar{\boldsymbol{u}}) + \cdots \end{aligned} \tag{33}$$

Therefore, neglecting high-order terms, we obtain

$$I_u \simeq f(\bar{\boldsymbol{u}}) + \frac{1}{2}\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{C}_u\} \tag{34}$$

where $\boldsymbol{H}$ is the Hessian evaluated at $\bar{\boldsymbol{u}}$. Consider now a complex-valued circular random vector $\boldsymbol{v}$ with mean $\bar{\boldsymbol{v}}$ and covariance matrix $\boldsymbol{C}_{\boldsymbol{v}}$, and assume we wish to compute

$$I_v = \int f(\boldsymbol{v})p(\boldsymbol{v})\, d\boldsymbol{v} \tag{35}$$

where $f(\cdot)$ is a real-valued scalar function of the complex-valued vector $\boldsymbol{v}$. Applying the result (34) to $\boldsymbol{u} = [\mathrm{Re}(\boldsymbol{v})^T\ \ \mathrm{Im}(\boldsymbol{v})^T]^T$ along with well-known results on the derivatives with respect to complex-valued parameters, it is straightforward to show that

$$I_v \simeq f(\bar{\boldsymbol{v}}) + \mathrm{Tr}\left\{\left.\frac{\partial^2 f}{\partial \boldsymbol{v}\partial \boldsymbol{v}^H}\right|_{\bar{\boldsymbol{v}}} \boldsymbol{C}_v\right\}. \tag{36}$$

The previous result is now used to compute $I_k = \int f_k(\boldsymbol{v}) p(\boldsymbol{v})\, d\boldsymbol{v}$ with

$$f_k(\boldsymbol{v}) = \frac{v_k}{1 + \sigma_s^2 \boldsymbol{v}^H \boldsymbol{C}^{-1} \boldsymbol{v}} \tag{37}$$

and where v_k is the kth component of $\boldsymbol{v}$. The first and second-order derivatives can be readily obtained as

$$\begin{aligned} \frac{\partial f_k}{\partial \boldsymbol{v}} &= \frac{-v_k \sigma_s^2 \boldsymbol{C}^{-1} \boldsymbol{v}}{\left(1 + \sigma_s^2 \boldsymbol{v}^H \boldsymbol{C}^{-1} \boldsymbol{v}\right)^2} \\ \frac{\partial^2 f_k}{\partial \boldsymbol{v} \partial \boldsymbol{v}^H} &= \frac{-\sigma_s^2 \left[v_k \boldsymbol{C}^{-1} + \boldsymbol{C}^{-1} \boldsymbol{v} \boldsymbol{e}_k^H\right]}{\left(1 + \sigma_s^2 \boldsymbol{v}^H \boldsymbol{C}^{-1} \boldsymbol{v}\right)^2} \\ &\quad + \frac{2 v_k \sigma_s^4 \boldsymbol{C}^{-1} \boldsymbol{v} \boldsymbol{v}^H \boldsymbol{C}^{-1}}{\left(1 + \sigma_s^2 \boldsymbol{v}^H \boldsymbol{C}^{-1} \boldsymbol{v}\right)^3} \end{aligned} \tag{38}$$

where $\boldsymbol{e}_k$ is a vector whose elements are all zero except the kth, which equals 1. Consequently

$$\begin{aligned} \operatorname{Tr}\left\{ \left. \frac{\partial^2 f_k}{\partial \boldsymbol{v} \partial \boldsymbol{v}^H} \right|_{\bar{\boldsymbol{v}}} \boldsymbol{C}_v \right\} &= \frac{-\sigma_s^2 \operatorname{Tr}\{\boldsymbol{C}^{-1} \boldsymbol{C}_v\} \bar{v}_k}{\left(1 + \sigma_s^2 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}\right)^2} \\ &+ \frac{2 \sigma_s^4 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \boldsymbol{C}_v \boldsymbol{C}^{-1} \bar{\boldsymbol{v}} \bar{v}_k}{\left(1 + \sigma_s^2 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}\right)^3} - \frac{\sigma_s^2 \boldsymbol{e}_k^H \boldsymbol{C}_v \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}}{\left(1 + \sigma_s^2 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}\right)^2} \end{aligned} \tag{39}$$

and hence

$$I_k \simeq \frac{\bar{v}_k}{1 + \sigma_s^2 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}} + \operatorname{Tr}\left\{ \left. \frac{\partial^2 f_k}{\partial \boldsymbol{v} \partial \boldsymbol{v}^H} \right|_{\bar{\boldsymbol{v}}} \boldsymbol{C}_v \right\}. \tag{40}$$

Gathering the previous equation for $k = 1, \ldots, m$ in a single vector results in

$$\int \frac{\boldsymbol{v}}{1 + \sigma_s^2 \boldsymbol{v}^H \boldsymbol{C}^{-1} \boldsymbol{v}} p(\boldsymbol{v})\, d\boldsymbol{v} \simeq \alpha \bar{\boldsymbol{v}} + \beta \boldsymbol{C}_v \boldsymbol{C}^{-1} \bar{\boldsymbol{v}} \tag{41}$$

with

$$\begin{aligned} \alpha &= \frac{-\sigma_s^2 \operatorname{Tr}\{\boldsymbol{C}^{-1} \boldsymbol{C}_v\}}{\left(1 + \sigma_s^2 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}\right)^2} + \frac{2 \sigma_s^4 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \boldsymbol{C}_v \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}}{\left(1 + \sigma_s^2 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}\right)^3} \\ \beta &= \frac{-\sigma_s^2}{\left(1 + \sigma_s^2 \bar{\boldsymbol{v}}^H \boldsymbol{C}^{-1} \bar{\boldsymbol{v}}\right)^2}. \end{aligned} \tag{42}$$

REFERENCES

[1] Y. Cailloce, G. Caille, B. Demolder, and G. Bekaert, "F.A.F.R. multi-beam antenna demonstrator for Ka-band multimedia via satellite mission," in *Proc. 52nd Int. Astronaut. Congr.*, Toulouse, France, Oct. 2001, pp. 1–5.

[2] A. Cherrette, R. Acosta, P. Lam, and S. Lee, "Compensation of reflector antenna surface distortion using an array feed," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 966–978, Aug. 1989.

[3] Y. Rahmat-Samii, "Array feeds for reflector surface distortion compensation: Concepts and implementation," *IEEE Antennas Propagat. Mag.*, vol. 32, pp. 20–26, Aug. 1990.

[4] H. V. Trees, *Optimum Array Processing*. New York: Wiley, 2002.

[5] A. Gershman, "Robust adaptive beamforming in sensor arrays," *Int. J. Electron. Commun.*, vol. 53, no. 6, pp. 305–314, Dec. 1999.

[6] B. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, pp. 397–401, July 1988.

[7] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1365–1376, Oct. 1987.

[8] M. Er and A. Cantoni, "An alternative formulation for an optimum beamformer with robustness capability," in *Proc. Inst. Elect. Eng. F, Radar, Sonar, Navigat.*, vol. 132, Oct. 1985, pp. 447–460.

[9] B. D. V. Veen, "Minimum variance beamforming with soft response constraints," *IEEE Trans. Signal Processing*, vol. 39, pp. 1964–1972, Sept. 1991.

[10] K. Bell and H. V. Trees, "Adaptive and nonadaptive beampattern control using quadratic beampattern constraints," in *Proc. 33rd Asilomar Conf.*, Pacific Grove, CA, Oct. 1999, pp. 486–490.

[11] R. Mailloux, "Covariance matrix augmentation to produce adaptive array pattern troughs," *Electron. Lett.*, vol. 31, no. 10, pp. 771–772, 1995.

[12] M. Zatman, "Production of adaptive array troughs by dispersion synthesis," *Electron. Lett.*, vol. 31, no. 25, pp. 2141–2142, 1995.

[13] J. Guerci, "Theory and application of covariance matrix tapers for robust adaptive beamforming," *IEEE Trans. Signal Processing*, vol. 47, pp. 977–985, Apr. 1999.

[14] M. Zatman, "Comments on theory and application of covariance matrix tapers for robust adaptive beamforming," *IEEE Trans. Signal Processing*, vol. 48, pp. 1796–1800, June 2000.

[15] D. Feldman and L. Griffiths, "A projection approach for robust adaptive beamforming," *IEEE Trans. Signal Processing*, vol. 42, pp. 867–876, Apr. 1994.

[16] J. Yu and C. Yeh, "Generalized eigenspace-based beamformers," *IEEE Trans. Signal Processing*, vol. 43, pp. 2453–2461, Nov. 1995.

[17] S. Vorobyov, A. Gershman, and Z. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, pp. 313–324, Feb. 2003.

[18] R. Lorenz and S. Boyd, "Robust minimum variance beamforming," in *Proc. 37th Asilomar Conf. Signals Syst. Comput.*, vol. 2, 2002, pp. 1345–1352, submitted for publication.

[19] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Processing*, vol. 51, pp. 1702–1715, July 2003.

[20] K. Bell, Y. Ephraim, and H. V. Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Processing*, vol. 48, pp. 386–398, Feb. 2000.

[21] A. Weiss and B. Friedlander, "'Almost blind' signal estimation using second-order moments," *Proc. Inst. Elect. Eng. F, Radar, Sonar Navigat.*, vol. 142, no. 5, pp. 213–217, Oct. 1995.

[22] ——, "Comparison of signal estimation using calibrated and uncalibrated arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 241–249, Jan. 1997.

[23] M. Viberg and A. Swindlehurst, "A Bayesian approach to auto-calibration for parametric array processing," *IEEE Trans. Signal Processing*, vol. 42, pp. 3495–3507, Dec. 1994.

[24] J. Yang and A. Swindlehurst, "The effects of array calibration errors on DF-based signal copy performance," *IEEE Trans. Signal Processing*, vol. 43, pp. 2724–2732, Nov. 1995.

[25] A. Swindlehurst, "A maximum a posteriori approach to beamforming in the presence of calibration errors," in *Proc. IEEE SSAP Conf*, Corfu, Greece, June 1996, pp. 82–85.

[26] B. Flanagan and K. Bell, "Array self-calibration with large sensor position errors," *Signal Process.*, vol. 81, no. 10, pp. 2201–2214, Oct. 2001.

[27] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

Olivier Besson (M'92) received the Ph.D. degree in signal processing in 1992 and the "Habilitation à Diriger des Recherches" degree in 1998, both from INPT, Tolousse, France.

He is currently an Associate Professor with the Department of Avionics and Systems, ENSICA, Toulouse. His research interests are in the general area of statistical signal and array processing with applications in communications and radar.

Dr. Besson served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from February 2000 to January 2003. He is a member of the IEEE SAM Technical Committee and the co-technical chairman of the IEEE SAM04 Workshop.

Andrei A. Monakov graduated from the Leningrad Institute of Aviation Instrument Making. Leningrad, U.S.S.R., in 1978. He received the Cand. Sci. degree in 1984 and the Dr. Sci. degree in 2000 from the St. Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia.

He is a Professor at the Radio Engineering Department, St. Petersburg State University of Aerospace Instrumentation. His current areas of interest include digital signal processing, radar theory, remote sensing, and air traffic control.

Christophe Chalus received the engineer degree in aerospace and the D.E.A. degree in signal processing, both from ENSICA, Toulouse, France, in 2002. He is currently pursuing the Ph.D. degree at the Cooperative Research Laboratory for Aeronautics and Space Telecommunications (TéSA), Toulouse. His thesis deals with beamforming problems with large arrays for multibeam satellite communications systems.